ORIGINAL RESEARCH ARTICLE

Optimizing Order of Administration for Concussion Baseline Assessment Among NCAA Student-Athletes and Military Cadets

Landon B. Lempke[1,2,3] · Robert C. Lynall[1] · Melissa N. Anderson[1] · Michael A. McCrea[4] · Thomas W. McAllister[5] · Steven P. Broglio[6] · Julianne D. Schmidt[1] · CARE Consortium Investigators

Abstract

Background Concussion pre-injury (i.e., baseline) assessments serve as a benchmark comparison point in the event an individual sustains a concussion and allows clinicians to compare to post-injury measures. However, baseline assessments must reflect the individual's true and most optimized performance to serve as a useful comparison. Mental fatigue and motivation throughout baseline testing may alter individual assessment performance, indicating an order of administration (OoA) may play an influential role in assessment outcomes.

Objective To examine the influence concussion baseline battery OoA has on symptom, postural stability, cognitive screening, and computerized neurocognitive test outcomes.

Methods We employed a retrospective observational cohort study to examine healthy collegiate student-athletes and military cadets ($n = 2898$, 19.0 ± 1.4 years, 66.1% male, 75.6% white, 54.4% Division-I) baseline assessment performance on the Sport Concussion Assessment Tool (SCAT; total symptom number and severity), Balance Error Scoring System (BESS; total error scores), Standardized Assessment of Concussion (SAC; total score), and Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT) domain scores (verbal and visual memory, visual-motor speed, reaction time). Assessments were binned to beginning, middle, or end tertiles based upon OoA. We used one-way ANOVAs with Tukey post-hoc t tests, 95% confidence intervals (CI), and Cohen's d effect sizes for significant models ($\alpha = 0.05$).

Results SCAT total symptom number (mean difference $= 2.23$; 95% CI 1.76–2.70; $d = 0.49$, $p < 0.001$) and severity (mean difference $= 5.58$; 95% CI 4.42–6.74; $d = 0.50$; $p < 0.001$) were lower when completed at the end of baseline testing compared to the middle. Total BESS errors were 1.06 lower when completed at the middle relative to the end (95% CI 0.43–1.69; $d = 0.17$; $p = 0.001$). Total SAC scores were better at the beginning relative to middle (mean difference $= 0.58$; 95% CI 0.25–0.90; $d = 0.33$; $p < 0.001$) and end (mean difference $= 0.44$; 95% CI 0.16–0.73; $d = 0.24$; $p = 0.001$). Verbal memory, visual memory, and reaction time performance were highest at the beginning ($p \leq 0.002$), while visual-motor speed performance was highest at the middle ($p = 0.001$).

Conclusion Completing baseline assessments in the order of (1) ImPACT, (2) SAC, (3) BESS, and (4) SCAT symptom checklist may improve performance across assessments collectively. Clinicians and researchers should consider completing baseline assessments in this order when possible to potentially aid in optimizing concussion baseline assessment performance and maximize post-concussion comparisons.

✉ Landon B. Lempke
lblempke@uga.edu

Extended author information available on the last page of the article

Key Points

Completing baseline assessments in the order of (1) ImPACT, (2) SAC, (3) BESS, and (4) SCAT symptom checklist may slightly improve performance across assessments jointly.

Completing baseline concussion assessments in this order is an easily implementable technique to potentially assist in improving baseline performance.

Clinicians and researchers should aim to reduce confounding factors to baseline performance where possible, with baseline order of administration being one of the numerous factors to control.

1 Introduction

Concussion is a complex and diffuse brain injury frequently occurring in sport across all levels of participation [1, 2]. Symptom, postural stability, and computerized neurocognitive testing assessments are key elements for optimal diagnosis and monitoring recovery [3, 4]. Each assessment serves a unique purpose in identifying the diverse post-concussion impairments, and when used in conjunction, these three assessment domains provide the strongest diagnostic accuracy to date [5–7]. Pre-injury baseline measures are frequently employed in sport [8] and military settings [9] to serve as a healthy comparison point in case of a future concussion [7, 10]. When used in this manner, it is imperative that baseline performance is optimized, but numerous confounders exist [11–14].

Previous work has identified several modifiable and non-modifiable factors that augment baseline concussion assessment performance. Factors such as group test setting [12], supervision [15], sleep [16], motivation [17], effort [13], and physical [18] and mental fatigue [19] can all be modified to maximize performance, while concussion history [11], biological sex [11], and age [20] cannot. Not accounting for modifiable factors may reduce peak effort and consequently produce a sub-optimal performance that can result in improper post-injury conclusions [13]. One unexplored and feasible consideration that could be modified to optimize baseline performance is the order of administration (OoA) assessments are applied. Yet, no information exists about how OoA might influence concussion baseline assessment performance.

The OoA is not standardized in concussion research and may result in sub-optimal test performance at baseline and post-injury. Randomizing OoA would likely have adverse effects as it would not eliminate or control for an OoA effect. Traditional neuropsychology work has examined and identified the OoA can greatly alter each assessment within a clinical battery [21–23], but has never been explored among concussion assessments. Determining a favorable concussion OoA may be advantageous for more stable outcome score performance and fewer invalid tests. Identifying the role of OoA may provide greater stability to assessment scores and optimal post-concussion comparison.

Mental fatigue is a commonly occurring phenomenon during the sustained cognitive engagement that negatively impacts task performance and grows in deleterious effect as time goes on [24, 25]. It is plausible that the time and complexity demand needed for thorough concussion baselines may inadvertently induce performance declines. For example, mental fatigue may disproportionately influence computerized neurocognitive performance or cognitive screening assessments due to sustained cognitive engagement. Similarly, symptom checklists presumably may also be prone to adverse effects after sustained cognitive activity due to checklists inquiring about drowsiness, fatigue, feeling slowed down, and difficulty concentrating or remembering. These cumulative speculations highlight the limited understanding of how baseline assessments may be adversely affected by OoA.

The purpose of our study was to examine the influence baseline concussion battery OoA has on symptom, postural stability, cognitive screening, and computerized neurocognitive test outcomes among collegiate student-athletes and military cadets. We hypothesized that (1) participants would perform worse on computerized neurocognitive testing, a cognitive screening assessment, and symptom reporting when administered later (i.e., middle or end) in the baseline battery compared to earlier (i.e., beginning or middle) in the battery and (2) postural stability would not differ based on battery order.

2 Methods

This study was part of the National Collegiate Athletic Association—Department of Defense Concussion Assessment, Research and Education (CARE) Consortium, a study aimed at examining the effects of concussion in collegiate student-athletes and the United States military cadets. We examined student-athletes and military cadets who completed previously described baseline testing [9] between Fall 2014 and Fall 2017. This study is reported in accordance with Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines [26]. The United States Army Medical Research and Materiel Command Human Research Protection Office, the University

of Michigan Institutional Review Board, and each CARE consortium sites' Institutional Review Board reviewed all study procedures. All participants provided written informed consent prior to participation.

2.1 Participants

The initial sample of 47,400 baseline cases (not individual participants) from collegiate student-athletes and military cadets were examined for eligibility, with 2898 individual baselines from one participant each meeting inclusion criteria (Table 1). A flowchart outlining sample exclusion is presented in Fig. 1. Participants were excluded if they self-reported a moderate or severe traumatic brain injury history, brain surgery history, English was not their primary language, or any disorder in the following categories: attentional deficit hyperactivity disorder (ADHD), learning disability, psychiatric, bipolar, depression, schizophrenic, balance, or sleep. Participants often completed a baseline assessment each year as part of the CARE Consortium, but only their first baseline assessment was included to ensure the independence of observations. Participants were included if they completed all baseline assessment components in a single testing session [9], had OoA assessment records completed, and computerized neurocognitive testing was completed and indicated as valid via the Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT) [9].

2.2 Baseline Assessments

The CARE Consortium concussion baseline assessments conducted have been previously described elsewhere [9]. In brief, baseline concussion assessments were conducted using current evidence-based [3] assessments: the Sport Concussion Assessment Tool 3rd edition (SCAT) symptom inventory, the Balance Error Scoring System (BESS), the Standardized Assessment of Concussion (SAC), and the Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT). The test–retest reliability of these assessments within these CARE Consortium data have been thoroughly described in a previous manuscript [27]. Additionally, a detailed demographic and health history case report form and the Brief Symptom Inventory (BSI)-18 were completed across all CARE Consortium sites [9], but are not reported or analyzed here as they are not yet commonly recommended or employed for concussion assessment in clinical practice [3, 8]. Therefore, a minimum of 6 assessments (SCAT, BESS, SAC, ImPACT, BSI-18, demographic, and health history form) were completed by each participant.

We examined the SCAT symptom inventory total symptom severity score and total symptom score [28, 29]. SCAT total symptom severity consisted of rating each of the 22

Table 1 Participant demographics (n = 2898)

	Frequency	(%)
Sex (male)	1917	66.1%
Concussion history (≥ 1)	545	18.8%
Race		
White	2192	75.6%
Black/African American	251	8.7%
Multiple races	224	7.7%
Asian	131	4.6%
Unknown/skipped	62	2.1%
Other	38	1.3%
Number of CARE consortium sites	27	90.0%
NCAA division		
Division I	1577	54.4%
Division II	979	33.8%
Division III	342	11.8%
Sport		
Military cadet—club, company athletics, intramural	721	24.9%
Military cadet—non-sport	377	13.0%
Football	352	12.1%
Cross country/track/field	302	10.4%
Soccer	153	5.3%
Swimming/diving	138	4.8%
Basketball	130	4.5%
Baseball	126	4.3%
Rowing/crew	93	3.2%
Softball	73	2.5%
Lacrosse	67	2.3%
Other	51	1.8%
Tennis	48	1.7%
Volleyball	47	1.6%
Water polo	42	1.4%
Ice hockey	40	1.4%
Field hockey	38	1.3%
Cheerleading	38	1.3%
Golf	31	1.1%
Wrestling	31	1.1%
	Mean	Std
Age (years)	18.95	1.39
Mass (kg)	75.88	16.22
Height (cm)	176.76	10.24
Concussion History[a]	1.33	0.68

CARE Concussion Assessment, Research and Education, *NCAA* National Collegiate Athletic Association

[a]Only includes student-athletes who reported ≥ 1 previous concussion

symptoms on a 6-point Likert scale (0–132 score range) with a higher score indicating greater symptom severity. The SCAT total symptom number consisted of counting each of the 22 symptoms listed if participants scored that
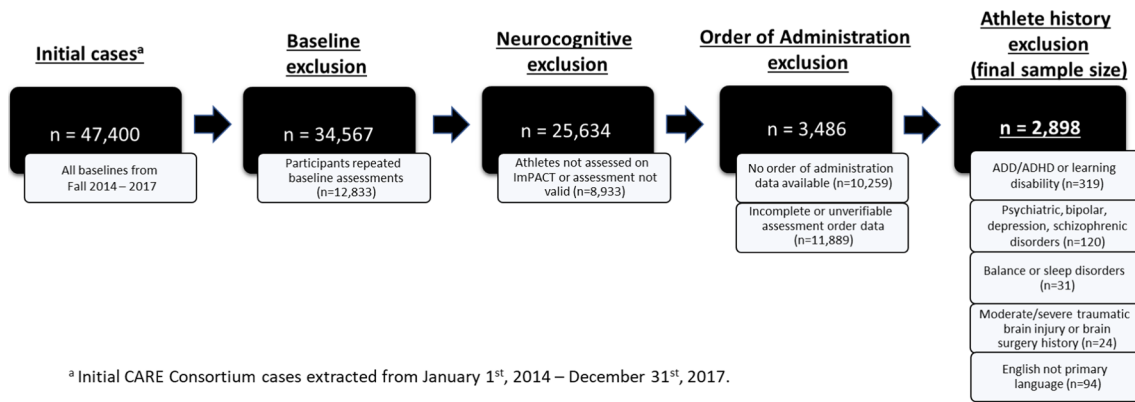
Fig. 1 Participant inclusion flowchart

symptom ≥ 1 (0–22 score range), with a higher score indicating greater symptom presence.

Participants completed the BESS assessment in three stances (double limb, single limb, and tandem limb) across two conditions (firm and foam surface) for 20 s each. Participants completed testing with their eyes closed and body in standardized positions. Deviations from the starting positions during trials were counted (max 10 each condition) as described in detail elsewhere [30]. The BESS total error scores were examined and calculated by summing the errors from all six trials, with a higher score indicating poorer balance performance (0–60 score range).

The SAC is a reliable and valid cognitive screening assessment frequently employed in sport [31, 32]. The SAC assesses the following cognitive domains: orientation (5-items), immediate memory (15-items), concentration (5-items), and delayed memory recall (5-items). A point is given for each item answered correctly, with a higher score indicating better performance. The SAC total score was examined (0–30 score range).

We examined ImPACT performance to examine general neurocognitive function since 90% (27/30) of CARE Consortium sites employed this reliable and valid assessment [27, 33]. The ImPACT composite scores (verbal memory, visual memory, visual-motor speed, and reaction time) were calculated as previously described [34], with higher composite scores indicating better performance, with the exception of reaction time. A higher reaction time composite score(s) indicates worse performance (i.e., more time needed to respond).

2.3 Baseline Assessment Order of Administration

No pre-determined testing order was mandated by the research protocol, resulting in numerous OoAs across CARE Consortium sites, as determined by the individual sites and test administrators. The OoA for all baseline assessments was recorded by the test administrator during the examination. All described baseline assessments, including the demographic and health history case report form and BSI-18, were utilized to determine the OoA. To ensure proper data entry and participant inclusion, only cases with complete OoA records were included. The OoA data were deemed unreliable, and therefore that case excluded if any baseline assessment OoA record was not entered or was greater than the max number of baseline assessments administrated (e.g., recorded as the 6th completed baseline assessment but only five assessments were completed), which indicated incomplete or inaccurate records. For example, if a participant completed five assessments (range = 5) but the BESS assessments corresponding OoA record (1–5) was missing or fell outside this range (< 1 or > 5), the participant was excluded. It is important to note the battery order was not an a priori CARE Consortium research aim, which resulted in a large number of cases being excluded due to incomplete or invalid baseline OoA provided (Fig. 1).

Assessments were categorized into beginning (0–33%), middle (34–66%) and end (67–100%) based upon when the individual assessment was administered that day and divided by the relative total number of assessments an individual completed to understand how OoA influences the outcome score. For example, completing the BESS as the second assessment out of five total, it would be classified as being completed in the middle (two divided by five; 40%). Binning was required due to some additional baseline assessments being employed in an inconsistent OoA across CARE sites [9], such as the vestibular ocular motor screening (VOMS) [35, 36], clinical reaction time measure [37, 38], or King-Devick test [39, 40]. These additional items however were completed among only 286 (9.9%) included participants. Tertile bins were employed to provide an understanding of OoA while maintaining statistical power across each level.

2.4 Statistical Analysis

Descriptive statistics for participant demographics and OoA were calculated. One-way ANOVA models were employed to examine how OoA (beginning, middle, end) influenced SCAT, BESS, SAC, and ImPACT outcomes, with Tukey adjusted post-hoc t-tests conducted to examine all group level differences. Sex was considered a priori as a covariate because sex distributions could differ between OoA groups and it was not feasible nor logical to exclude a certain sex. Preliminary chi-square tests were not statistically different for sex and administration frequencies for baseline assessment ($p \geq 0.137$) and therefore was not included in the ANOVA models. Some assessments OoA levels had insufficient sample size ($n < 100$) to allow optimal statistical power for group comparisons and were therefore dropped from analysis and the two remaining levels were examined using independent t tests. Cohen's d effect sizes and mean difference 95% confidence intervals (CI) were calculated between each assessment group level. Cohen's d effect sizes were interpreted according to conventional statistical guides with ≤ 0.20, 0.21–0.79, and ≥ 0.80 used as thresholds for small, medium, and large magnitude effects, respectively [41]. Linearity, normality, equal variance, and influential outliers were assessed, with no violations present. All statistical analyses were conducted using the R Project for Statistical Programming (v 3.4.3, Murray Hills, NJ, USA) [42] and $\alpha = 0.05$ a priori.

3 Results

A total of 2,898 participants across 27 of 30 (90%) of participating CARE Consortium sites met inclusion criteria. Participant demographics are presented in Table 1. The frequencies and percentages of baseline assessments across OoA are presented in Table 2, and all means, standard deviations, and statistical model parameters are presented in Table 3.

Table 2 Baseline order of administration across assessments

	Beginning	Middle	End
SCAT symptoms	32 (1.1%)[a]	2445 (84.4%)	421 (14.5%)
BESS	71 (2.4%)[a]	459 (15.9%)	2368 (81.7%)
SAC	264 (9.1%)	577 (19.9%)	2057 (71.0%)
ImPACT	1947 (67.2%)	295 (10.2%)	656 (22.6%)

SCAT Sport Concussion Assessment Tool, *BESS* Balance Error Scoring System, *SAC* Standardized Assessment of Concussion, ImPACT Immediate Post-Concussion Assessment and Cognitive Testing

[a]Level dropped from statistical analysis due to small sample size

3.1 SCAT Symptom Checklist Order of Administration

Only 32 (1.1%) participants completed the SCAT symptom checklist at the beginning OoA level and were excluded from the statistical model, but descriptive statistics are reported in Table 3. The SCAT total symptom number and total symptom severity were examined via independent t tests. Total symptom number was significantly different between OoA ($p < 0.001$), with SCAT total symptom decreasing by 2.23 (95% CI 1.76–2.70; $d = 0.49$) with a medium effect size when completed at the end of baseline testing compared to the middle (Table 3 and Fig. 2A). Total symptom severity was similar with significantly decreased severities at the end of baseline testing relative to the middle (mean difference = 5.58; 95% CI 4.42–6.74; $d = 0.50$; $p < 0.001$; Fig. 2B).

3.2 BESS Order of Administration

Only 71 (2.4%) participants completed BESS at the beginning OoA level and were excluded from the statistical model, but descriptive statistics are reported in Table 3. The BESS total error scores were therefore assessed via an independent t test. Total BESS errors were significantly higher ($p = 0.001$), with participants completing BESS at the end of the battery committing 1.06 more total errors with a small effect size (95% CI 0.43–1.69; $d = 0.17$) when compared to those completing BESS in the middle (Table 3 and Fig. 2C).

3.3 SAC Order of Administration

Total SAC score was significantly different between beginning, middle, and end OoA groups ($p < 0.001$; Table 3). Beginning group SAC scores were statistically greater than middle (mean difference = 0.58; 95% CI 0.25–0.90; $d = 0.33$; $p < 0.001$) and end (mean difference = 0.44; 95% CI 0.16–0.73; $d = 0.24$; $p = 0.001$) with medium effect sizes. Total SAC score was not statistically different between middle and end OoA groups ($p = 0.278$; Fig. 2D).

3.4 ImPACT Order of Administration

Significant OoA differences were present for all four ImPACT domain scores ($p \leq 0.002$; Table 3) and are presented in Fig. 3. The verbal memory domain score was greater when taken in the beginning than middle (mean difference = 2.63; 95% CI 1.22–4.04; $d = 0.28$; $p < 0.001$) and end (mean difference = 1.39; 95% CI 0.37–2.41; $d = 0.15$; $p = 0.004$) OoAs with medium and small effect

Table 3 Baseline order of administration ANOVA models

	Beginning Mean (SD)	Middle Mean (SD)	End Mean (SD)	p value[b] (t-statistic or F ratio)
SCAT symptom number	2.44[c] (3.45)[c]	4.34 (4.74)	2.11 (3.02)[a]	< 0.001 $T_{2864} = 9.32$
SCAT symptom severity	3.84[c] (7.02)[c]	8.76 (11.92)	3.18 (5.39)[a]	< 0.001 $T_{2864} = 9.44$
BESS total errors	11.87[c] (4.55)[c]	13.27 (6.12)[a]	14.33 (6.36)	0.001 $T_{2825} = -3.28$
SAC total score	27.85 (1.67)[a]	27.27 (1.75)	27.40 (1.90)	< 0.001 $F_{2, 2895} = 8.94$
ImPACT[d]				
Verbal memory	89.54 (9.23)[a]	86.91 (11.37)	88.15 (9.99)	< 0.001 $F_{2, 2895} = 12.49$
Visual memory	80.75 (12.14)[a]	78.51 (13.05)	77.86 (13.14)	< 0.001 $F_{2, 2895} = 15.14$
Visual motor speed	41.42 (6.16)	42.63 (6.66)[a]	41.06 (6.72)	0.002 $F_{2, 2895} = 6.38$
Reaction time	0.59 (0.08)	0.58 (0.09)[a]	0.60 (0.09)	0.002 $F_{2, 2895} = 6.01$

[a]Indicate the best performance statistically within a given baseline assessment outcome score

[b]p values from either omnibus one-way ANOVA models or independent t tests

[c]Beginning group dropped from statistical analysis due to small sample size (Table 2), and therefore the p value and t statistic reflect the middle and end group comparison

[d]ImPACT domain scores presented as standard composite scores

SCAT Sport Concussion Assessment Tool, *BESS* Balance Error Scoring System, *SAC* Standardized Assessment of Concussion, *ImPACT* Immediate Post-Concussion Assessment and Cognitive Testing

sizes, respectively. The verbal memory domain score was not statistically different between middle and end OoA groups ($p = 0.158$; Fig. 3A).

The visual memory domain score was better among the beginning OoA group with a small effect size compared to the middle (mean difference = 2.24; 95% CI 0.41–4.07; $d = 0.18$; $p = 0.011$) and medium effect size compared to the end (mean difference = 2.89; 95% CI 1.57–4.21; $d = 0.23$; $p < 0.001$) groups. The visual memory domain score was not statistically different between middle and end OoA groups ($p = 0.735$; Fig. 3B).

The visual motor speed domain score was statistically different, but inconsistent, across group comparisons. The beginning OoA group was lower with a medium effect compared to the middle (mean difference = − 1.21; 95% CI − 2.15 to − 0.29; $d = -0.19$; $p = 0.006$) but was not different compared to the end group ($p = 0.436$). Visual motor speed was higher among middle group administration compared to the end group with a medium effect size (mean difference = 1.57; 95% CI 0.53–2.61; $d = 0.23$; $p = 0.001$; Fig. 3C).

The reaction time domain score model was statistically different overall, but post-hoc t tests demonstrated only the beginning OoA group was faster with a small effect size compared to the end group (mean difference = − 0.01; 95% CI − 0.00 to − 0.02; $d = -0.13$; $p = 0.008$). No other

OoA comparisons were statistically significant ($p \geq 0.052$; Fig. 3D).

4 Discussion

Our findings indicate that symptom number and severity, balance, cognitive screening assessment, and computerized neurocognitive assessment performance are significantly altered based upon when they are completed during baseline testing. Though statistically significant, the immediate clinical meaning of the differences across each outcome is not clear due to varying effect size magnitudes and mean differences. Our findings however still collectively point towards a potentially optimal baseline OoA in hopes of reducing measurement error with minimal effort to researchers and clinicians.

Numerous confounders affecting optimal baseline assessment performance have been established [11, 12, 27, 43], highlighting specific factors clinicians and researchers should account for when possible. Our findings add to existing baseline assessment considerations but simultaneously provide a relatively simple solution. For example, the SCAT symptom checklist resulted in considerable symptom number and severity differences based on OoA. Though there
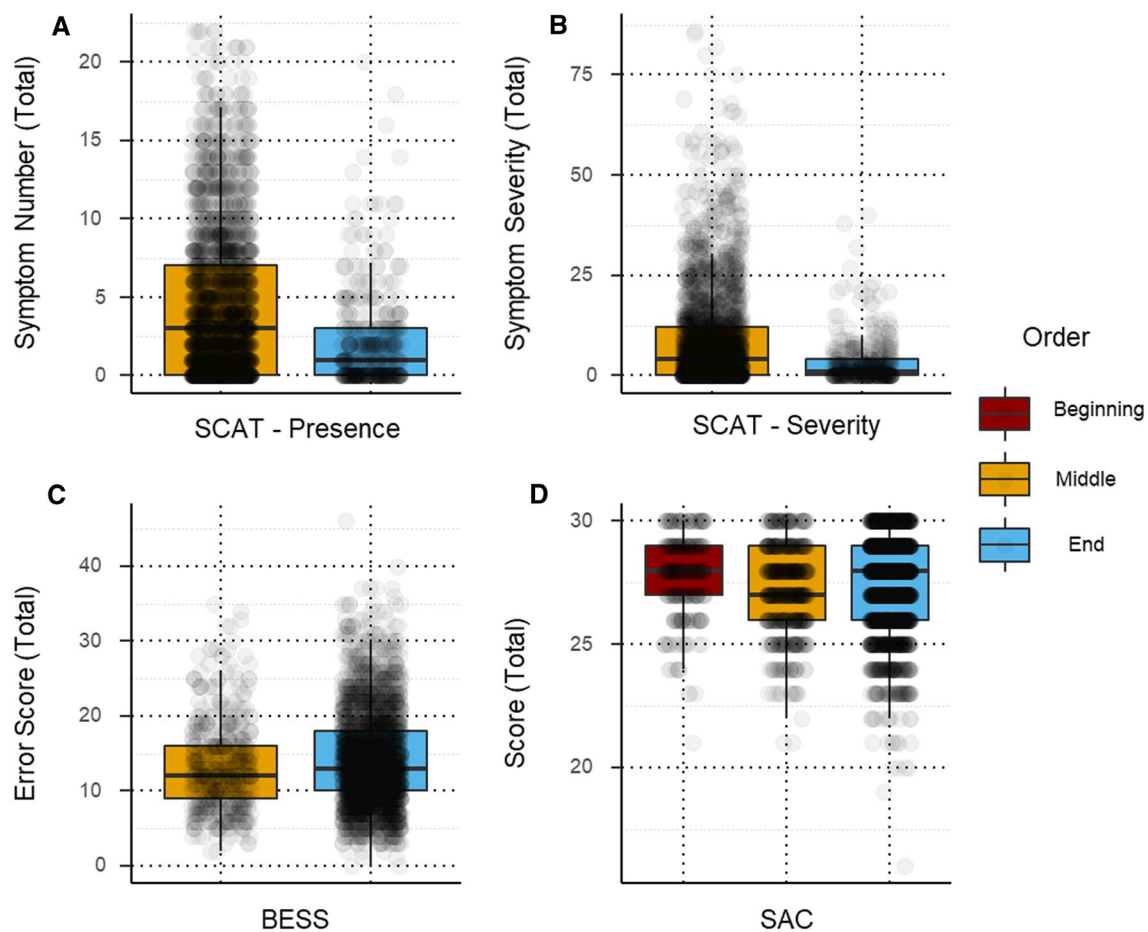
Fig. 2 **a** SCAT total symptom number, **b** SCAT total symptom severity, **c** BESS total error score, and **d** SAC total score across beginning (red), middle (yellow), and end (blue) order of administration groupings. The SAC total score y-axis is truncated to more clearly show all participant datapoints. The beginning categories are not displayed for SCAT or BESS due to small sample sizes leading to this category being excluded from statistical analysis. Each participant's outcome score for each assessment is plotted via jittered, 95% transparent dots. The thick, black line in box plots represents median values, box widths represent first (25%) and third (75%) quantiles, and whiskers represent box quantiles $\pm 1.5 \times$ interquartile range. *SCAT* Sport Concussion Assessment Tool, *BESS* Balance Error Scoring System, *SAC* Standardized Assessment of Concussion

were not enough participants completing SCAT at the beginning of baseline testing to statistically assess, we observed that completing it at the end versus the middle resulted in symptom number and severity decreasing by 2.23 and 5.58, respectively. The observed symptom differences may be clinically meaningful as moderate effect sizes ($d = 0.49$ and 0.50) were identified, and mean differences approach the previously published [27] 90% non-parametric change score confidence interval (90% ΔCI) values of 3 and 6 for total symptom number and severity, respectively. The 90% ΔCI values provide researchers and clinicians a degree of certainty (i.e. 90%) for identifying if outcome scores have changed when an assessment has improved or declined and are similar in interpretation to a reliable change index but differ in their mathematical calculation [27, 44, 45]. The 90% ΔCI provides direct clinical insights to if an individual's or a cohort's outcome score has changed, and are

discussed throughout to provide immediate clinical relevance to our findings.

The clinical implications of our findings suggest the SCAT symptom checklist may be better to implement at the end of baseline assessments rather than at the middle of the testing battery. Our findings contradict our hypothesis that the potential cognitive fatigue induced during baseline assessments would exacerbate symptoms when completed at the end. Though we did not statistically compare the SCAT symptom performance beginning OoA category with middle or end categories due to the small sample size, the reported group estimates for all three categories (Table 3) still suggest completing the SCAT symptom checklist at the end is optimal. Cognitive fatigue is an established phenomenon [24, 25], but it is possible it is not applicable to symptom assessments or even all baseline assessments, where there are often short pauses between assessments. Baseline assessments
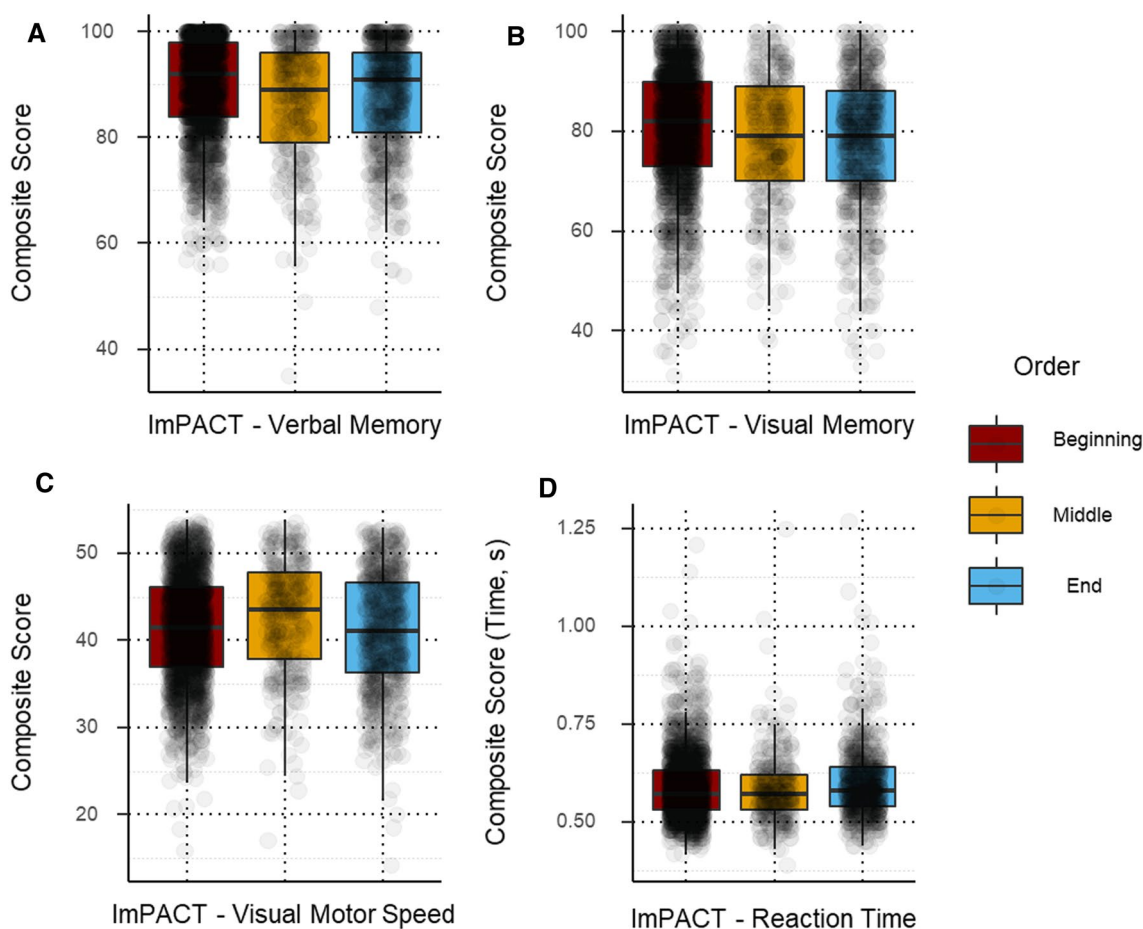
Fig. 3 ImPACT domain scores of **a** verbal memory, **b** visual memory, **c** visual motor speed, and **d** reaction time across beginning (red), middle (yellow), and end (blue) order of administration groupings. All ImPACT domain score *y* axes are truncated to more clearly show all participant datapoints. Each participant's outcome score for each assessment is plotted via jittered, 95% transparent dots. The thick, black line in box plots represents median values, box widths represent first (25%) and third (75%) quantiles, and whiskers represent box quantiles ± 1.5 × interquartile range. *ImPACT* Immediate Post-Concussion Assessment and Cognitive Testing

anecdotally can be a stressor for student-athletes given the cognitive and motor challenges employed and the sports environment sometimes putting arbitrary pressure on them to "pass their baseline". It is possible completing the symptom checklist at the end may better capture an individual's true pre-injury status and serve as a better comparator post-injury. It is important to note that the ImPACT has a symptom checklist embedded at the beginning of testing and may confound our findings. This may also be a limitation due to serial symptom assessment, but previous findings indicate healthy participants do not demonstrate meaningful changes in symptom number or severity across the daily assessment [46].

There was not a large enough sample size completing BESS, in the beginning, to fully examine across OoA and limits our insights. Student-athletes on average committed 1.06 more total errors on BESS when administered at the end of baseline testing relative to the middle. Our findings supported our hypothesis that later BESS baseline testing would result in worse performance. Balance assessments may be adversely altered due to cognitive fatigue being reported to alter both physiological and perceived physical performance capabilities [47], and may partially explain this finding. However, this statistical difference is likely not clinically meaningful due to the small magnitude effect size ($d = 0.17$) and total mean error difference only being around 1. The 90%ΔCI for BESS total errors ranges from 5 to 6 [27], suggesting OoA is a minimal consideration on its own. Other factors such as inter- and intra-tester reliability [43], sex [48], age [48], and physical exertion prior to assessment[49] play a larger role in altering BESS performance. However, it is possible OoA has an additive effect on errors when coupled with established influential factors. Given the minimal effort required for switching assessment OoA, clinicians could consider completing

BESS during the middle of baseline testing when possible to potentially reduce measurement variability.

The SAC total score was statistically better when completed at the beginning of the baseline assessment relative to the middle or end (mean difference = 0.58 and 0.44; Cohen's d = 0.33 and 0.24, respectively). Our findings supported our hypothesis that earlier baseline order completion would result in better performance. Cognitive fatigue may be applicable to baseline concussion assessments such as SAC, but likely not in a clinically meaningful way. These statistically significant findings likely have minimal clinical implications on their own as the mean differences are far from the 2-point 90% ΔCI and the medium effect sizes are relatively low. The mean difference values observed in this study are similar to previous work identifying males and those with ADHD have worse performance than females and those without ADHD by 0.50 and 0.71, respectively [50]. Our findings suggest SAC performance is minimally influenced by OoA. Clinicians and researchers may not need to consider when SAC is completed in the baseline OoA.

The ImPACT platform is utilized by 84% of sports medicine healthcare providers in the United States, making it the most widely implemented computerized neurocognitive test [8]. Numerous confounders to optimal baseline ImPACT outcome scores exist and mark the importance of controlling for variability when possible [11–14]. Our findings indicate each ImPACT domain was influenced by OoA, but similar to SAC, these differences are unlikely to be clinically important. Verbal memory, visual memory, and reaction time resulted in better outcome scores when completed at the beginning of baseline testing relative to middle or end assessments, but visual-motor speed performance was minimally worse when completed at the beginning. Our hypothesis was supported by our findings given the better ImPACT performance when completed earlier during baseline testing. Similar to SCAT and SAC though, cognitive fatigue may be a negligible contributor to the observed findings. The ImPACT does require sustained attention and effort, but the subtests employed change to assess the four domain scores and may help alleviate cognitive fatigue. Though three of the four domain scores were better when completed at the beginning, these statistical improvements possessed marginal mean differences. The published 90% ΔCIs for verbal memory (9.6), visual memory (13.6), visual-motor speed (4.4), and reaction time (0.08 s) are 4.6–8.0 times greater than the mean differences between beginning and end orders observed [27]. The previously established influential factors, such as sleep [16], motivation,[17], effort [13], or physical[18] and mental fatigue[19], may have an augmented effect with OoA that was not explored in this study. Given the minimal additional effort needed to alter baseline OoA, clinicians and researchers could consider implementing ImPACT at the beginning to potentially optimize its assessment performance across the majority of domains.

Though the observed OoA mean differences and effect sizes observed among individual baseline assessments may indicate minimal performance changes, making clinical changes where possible to reduce variability and optimize assessment performance may be beneficial. Our findings potentially point towards an improved baseline OoA when considering all assessments together. Both the SCAT symptom number and severity score were the most affected baseline measure relative to the 90% ΔCI. Symptom checklists have consistently been the strongest diagnostic tool for concussion assessment [5–7], and our findings potentially indicate completing them at the end of the pre-injury baseline may provide optimal order. The majority of ImPACT domain scores were better during the beginning of baseline testing, and BESS total errors were lowest when completed in the middle. The SAC assessment has statistically favored the beginning of OoA, but its mean difference relative to the 90%ΔCI was the least considerable among all baseline assessments. Therefore, we propose an optimal order for completing baseline assessments that may be: (1) ImPACT, (2) SAC, (3) BESS, and (4) SCAT symptom checklist. Given the SCAT5 [51] is comprised of all these assessments except ImPACT, the SCAT5 and future renditions could theoretically benefit from the OoA being reordered to SAC, BESS, and SCAT symptom checklist to potentially optimize this standardized assessments performance. However, reorganization of the SCAT5 should be further studied and explored to ensure its validity is maintained. Clinicians and researchers could consider implementing baseline assessments in this order to reduce assessment error and provide more accurate insights to pre-injury status. This optimal baseline order may also potentially be implemented with post-injury assessments to minimize assessment error. Given OoA altered all pre-injury assessment outcomes, it is plausible OoA will augment post-injury performance as well. Researchers should aim to fully explore if and to what degree OoA alters post-injury assessments to establish if the proposed order still holds true.

4.1 Limitations

Our study examined a collegiate athletic and military cadet sample and these findings may not be representable or applicable to other populations. Our study was a retrospective observational study design with inherent limitations such as unequal and uncontrolled OoA sample size that contributed to the small sample sizes completing the SCAT symptom checklist and BESS in the beginning OoA. Because negligible sample sizes were present for beginning OoA SCAT symptom checklist and BESS, we were unable to statistically assess if beginning OoA produced better outcomes.

Additionally, we binned baseline OoA to account for variation across CARE sites, which can indirectly influence overall findings. Some CARE Consortium sites also conducted supplemental assessments beyond those discussed [9], which accounted for 286 participants (9.9%) of cases presented. Though the majority of supplemental assessments take minimal time to complete, these short additional assessments may confound our findings. Our study however provides preliminary insights into how OoA may influence baseline assessments using previously collected data. Future research should examine OoA using a prospective, and more controlled study design with randomization to fully understand the role OoA plays on concussion baseline performance.

5 Conclusions

Concussion baseline assessments consisting of SCAT symptom checklist, SAC, BESS, and ImPACT are all influenced by the OoA they are each completed in. Not all assessments resulted in clinically meaningful differences based on OoA alone. Completing the SCAT symptom checklist at the end of baseline testing resulted in less total symptom number and severity. The BESS assessment produced fewer total errors when completed during the middle of baseline testing, SAC favored the beginning of baseline testing, and ImPACT verbal memory, visual memory, and reaction time were better at the beginning of baseline testing. Our results may indicate completing baseline assessments in the order of (1) ImPACT, (2) SAC, (3) BESS, and (4) SCAT symptom checklist could slightly optimize performance across assessments collectively, though future prospective research is warranted. Clinicians may consider completing baseline assessments in this order to potentially reduce assessment error and provide more accurate insights to pre-injury status. Our findings also may indicate the SCAT5 and future versions could be reorganized to (1) SAC, (2) BESS, and (3) SCAT symptom checklist to potentially reduce assessment variability, though future research is warranted.

Declarations

References

1. Guskiewicz KM, Weaver NL, Padua DA, Garrett WE. Epidemiology of concussion in collegiate and high school football players. Am J Sports Med. 2000;28:643–50.

2. Halstead ME, Walter KD. The Council on sports medicine and fitness sport-related concussion in children and adolescents. Pediatrics. 2010;126:597–615.

3. McCrory P, Meeuwisse W, Dvorak J, Aubry M, Bailes J, Broglio S, et al. Consensus statement on concussion in sport—the 5th international conference on concussion in sport held in Berlin, October 2016. Br J Sports Med. 2017;bjsports-2017-097699.

4. Broglio SP, Puetz TW. The effect of sport concussion on neurocognitive function, self-report symptoms and postural control. Sports Med. 2008;38:53–67.

5. Broglio SP, Macciocchi SN, Ferrara MS. Sensitivity of the concussion assessment battery. Neurosurgery. 2007;60:1050–7 (**discussion 1057–1058**).

6. Resch JE, Brown CN, Schmidt J, Macciocchi SN, Blueitt D, Cullum CM, et al. The sensitivity and specificity of clinical measures of sport concussion: three tests are better than one. BMJ Open Sport Exerc Med. 2016;2:e000012.

7. Broglio SP, Harezlak J, Katz B, Zhao S, McAllister T, McCrea M, et al. Acute sport concussion assessment optimization: a prospective assessment from the CARE consortium. Sports Med [Internet]. 2019 [cited 2019 Jul 28]. https://doi.org/10.1007/s40279-019-01155-0.

8. Lempke LB, Schmidt JD, Lynall RC. Athletic trainers' concussion-assessment and -management practices: an update. J Athl Train. 2020;51:17–26.

9. Broglio SP, McCrea M, McAllister T, Harezlak J, Katz B, Hack D, et al. A National Study on the Effects of Concussion in Collegiate Athletes and US Military Service Academy Members: the NCAA–DoD Concussion Assessment, Research and Education (CARE) Consortium Structure and Methods. Sports Med [Internet]. 2017 [cited 2017 Mar 24]. https://doi.org/10.1007/s40279-017-0707-1.

10. Louey AG, Cromer JA, Schembri AJ, Darby DG, Maruff P, Makdissi M, et al. Detecting cognitive impairment after concussion: sensitivity of change from baseline and normative data methods using the CogSport/Axon cognitive test battery. Arch Clin Neuropsychol. 2014;29:432–41.

11. Covassin T, Elbin R, Kontos A, Larson E. Investigating baseline neurocognitive performance between male and female athletes with a history of multiple concussion. J Neurol Neurosurg Psychiatry. 2010;81:597–601.

12. Moser RS, Schatz P, Neidzwski K, Ott SD. Group versus individual administration affects baseline neurocognitive test performance. Am J Sports Med. 2011;39:2325–30.

13. Anderson MN, Lempke LB, Bell DH, Lynall RC, Schmidt JD. The ability of CNS vital signs to detect coached sandbagging performance during concussion baseline testing: a randomized control trial. Brain Inj. 2020;34:369–74.

14. Walton SR, Broshek DK, Freeman JR, Hertel J, Meyer JP, Erdman NK, et al. Institutionally based ImPACT Test® normative values may differ from manufacturer-provided normative values. Arch Clin Neuropsychol. 2020;35:275–82.

15. Kuhn AW, Solomon GS. Supervision and computerized neurocognitive baseline test performance in high school athletes: an initial investigation. J Athl Train. 2014;49:800–5.

16. Sufrinko A, Pearce K, Elbin RJ, Covassin T, Johnson E, Collins M, et al. The effect of preinjury sleep difficulties on neurocognitive impairment and symptoms after sport-related concussion. Am J Sports Med. 2015;43:830–8.

17. Bailey CM, Echemendia RJ, Arnett PA. The impact of motivation on neuropsychological performance in sports-related mild traumatic brain injury. J Int Neuropsychol Soc [Internet]. 2006 [cited 2019 Feb 4];12. http://www.journals.cambridge.org/abstract_S1355617706060619.

18. Covassin T, Weiss L, Powell J, Womack C. Effects of a maximal exercise test on neurocognitive function. Yearb Sports Med. 2008;2008:10–1.

19. Boksem MAS, Tops M. Mental fatigue: costs and benefits. Brain Res Rev. 2008;59:125–39.

20. Register-Mihalik JK, Kontos DL, Guskiewicz KM, Mihalik JP, Conder R, Shields EW. Age-related differences and reliability on computerized and paper-and-pencil neurocognitive assessment batteries. J Athl Train. 2012;47:297–305.

21. Lloyd KP, Higginson CI, Lating JM, Coiro MJ. Test-order effects: a measure of verbal memory influences performance on a measure of verbal fluency. Appl Neuropsychol Adult. 2012;19:299–304.

22. Brandon AD, Chavez EL. Order and delay effects on neuropsychological test presentation: the Halstead Category and Wisconsin Card Sorting Tests. Int J Clin Neuropsychol. 1985;7:152–3.

23. Franzen MD, Smith SS, Paul DS, Maclnnes WD. Order effects in the administration of the booklet category test and Wisconsin Card Sorting Test. Arch Clin Neuropsychol. 1993;8:105–10.

24. Ackerman PL, Kanfer R. Test length and cognitive fatigue: an empirical examination of effects on performance and test-taker reactions. J Exp Psychol Appl. 2009;15:163–81.

25. Boksem MAS, Meijman TF, Lorist MM. Effects of mental fatigue on attention: an ERP study. Cogn Brain Res. 2005;25:107–16.

26. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008;61:344–9.

27. Broglio SP, Katz BP, Zhao S, McCrea M, McAllister T, CARE Consortium Investigators. Test-retest reliability and interpretation of common concussion assessment tools: findings from the NCAA-DoD CARE consortium. Sports Med Auckl NZ. 2017;48:1255–68.

28. Medicine BPGL and BA of S and E. Sport Concussion Assessment Tool—3rd Edition. Br J Sports Med. 2013;47:259.

29. Guskiewicz KM, Register-Mihalik J, McCrory P, McCrea M, Johnston K, Makdissi M, et al. Evidence-based approach to revising the SCAT2: introducing the SCAT3. Br J Sports Med. 2013;47:289–93.

30. Guskiewicz KM, Ross SE, Marshall SW. Postural stability and neuropsychological deficits after concussion in collegiate athletes. J Athl Train. 2001;36:263–73.

31. McCrea M, Kelly JP, Kluge J, Ackley B, Randolph C. Standardized Assessment of Concussion in football players. Neurology. 1997;48:586–8.

32. McCrea M, Kelly JP, Randolph C, Kluge J, Bartolic E, Finn G, et al. Standardized Assessment of Concussion (SAC): On-Site Mental Status Evaluation of the Athlete: J Head Trauma Rehabil. 1998;13:27–35.

33. Resch JE, Schneider MW, Munro CC. The test-retest reliability of three computerized neurocognitive tests used in the assessment of sport concussion. Int J Psychophysiol. 2018;132:31–8.

34. Schatz P, Pardini J, Lovell M, Collins M, Podell K. Sensitivity and specificity of the ImPACT Test Battery for concussion in athletes. Arch Clin Neuropsychol. 2006;21:91–9.

35. Mucha A, Collins MW, Elbin RJ, Furman JM, Troutman-Enseki C, DeWolf RM, et al. A brief vestibular/ocular motor screening (VOMS) assessment to evaluate concussions: preliminary findings. Am J Sports Med. 2014;42:2479–86.

36. Sufrinko AM, Mucha A, Covassin T, Marchetti G, Elbin RJ, Collins MW, et al. Sex differences in vestibular/ocular and neurocognitive outcomes after sport-related concussion. Clin J Sport Med. 2017;27:133–8.

37. Caccese JB, Eckner JT, Franco-MacKendrick L, Hazzard JB, Ni M, Broglio SP, et al. Clinical reaction-time performance factors in healthy collegiate athletes. J Athl Train. 2020;55:601–7.

38. Lempke LB, Howell DR, Eckner JT, Lynall RC. Examination of reaction time deficits following concussion: a systematic review and meta-analysis. Sports Med. 2020;50:1341–59.

39. Galetta KM, Brandes LE, Maki K, Dziemianowicz MS, Laudano E, Allen M, et al. The King-Devick test and sports-related concussion: study of a rapid visual screening tool in a collegiate cohort. J Neurol Sci. 2011;309:34–9.

40. Tjarks BJ, Dorman JC, Valentine VD, Munce TA, Thompson PA, Kindt SL, et al. Comparison and utility of King-Devick and ImPACT(R) composite scores in adolescent concussion patients. J Neurol Sci. 2013;334:148–53.

41. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale: Erlbaum Associates; 1988.

42. RStudio Team. RStudio: integrated development environment for R [Internet]. Boston, MA: RStudio, Inc.; 2016. Available from: http://www.rstudio.com/.

43. Bell DR, Guskiewicz KM, Clark MA, Padua DA. Systematic review of the balance error scoring system. Sports Health. 2011;3:287–95.

44. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol. 1991;59:12–9.

45. Van der Vaart AW. Asymptotic statistics. Cambridge: Cambridge University Press; 2000.

46. Balasundaram AP, Athens J, Schneiders AG, McCrory P, Sullivan SJ. Day-to-day variability of post-concussion-like symptoms reported over time by a non-concussed cohort. Brain Inj. 2016;30:1599–604.

47. Marcora SM, Staiano W, Manning V. Mental fatigue impairs physical performance in humans. J Appl Physiol. 2009;106:857–64.

48. Ozinga SJ, Linder SM, Koop MM, Dey T, Figler R, Russman AN, et al. Normative performance on the balance error scoring system by youth, high school, and collegiate athletes. J Athl Train. 2018;53:636–45.

49. Wilkins JC, Valovich McLeod TC, Perrin DH, Gansneder BM. Performance on the balance error scoring system decreases after fatigue. J Athl Train. 2004;39:156–61.

50. Chin EY, Nelson LD, Barr WB, McCrory P, McCrea MA. Reliability and validity of the sport concussion assessment tool-3 (SCAT3) in high school and collegiate athletes. Am J Sports Med. 2016;44:2276–85.

51. Echemendia RJ, Meeuwisse W, McCrory P, Davis GA, Putukian M, Leddy J, et al. The Sport Concussion Assessment Tool 5th edition (SCAT5): background and rationale. Br J Sports Med. 2017;51:848–50.

Authors and Affiliations

Landon B. Lempke[1,2,3] [ORCID] · **Robert C. Lynall[1]** · **Melissa N. Anderson[1]** · **Michael A. McCrea[4]** · **Thomas W. McAllister[5]** · **Steven P. Broglio[6]** · **Julianne D. Schmidt[1]** · **CARE Consortium Investigators**

1 UGA Concussion Research Laboratory, Department of Kinesiology, University of Georgia, Ramsey Student Center, 330 River Rd., Athens, GA 30606, USA

2 Division of Sports Medicine, Boston Children's Hospital, Boston, MA, USA

3 Micheli Center for Sports Injury Prevention, Waltham, MA, USA

4 Department of Neurosurgery and Neurology, Milwaukee, WI, USA

5 Department of Psychiatry, School of Medicine, Indianapolis, IN, USA

6 Michigan Concussion Center, University of Michigan, Ann Arbor, MI, USA